

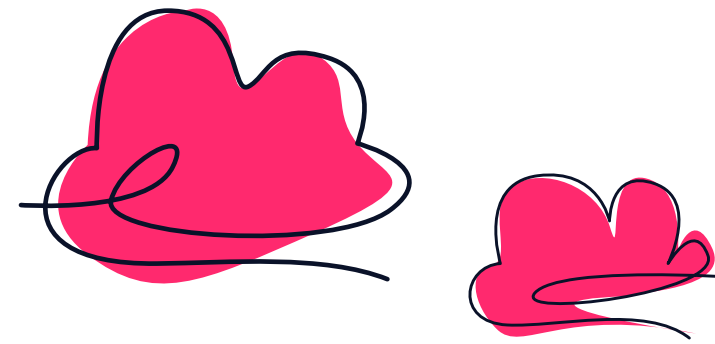


# The Art of Right-Sizing Your Cloud

How to Prevent Overprovisioning  
and Only Pay For the Cloud  
Resources You Need



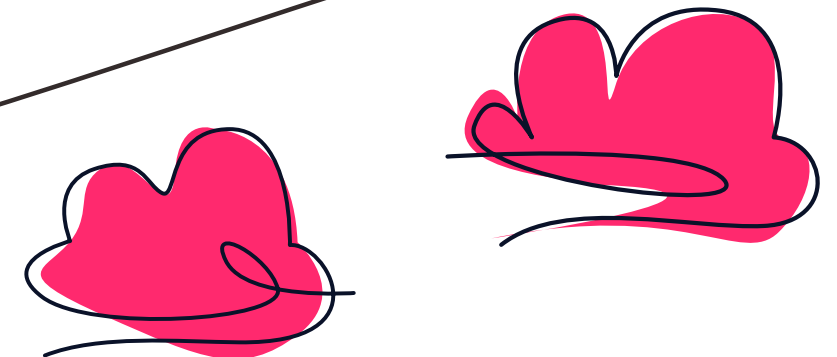
Determining the number of cloud resources that your organization needs is a balancing act. On the one hand, if you don't budget for enough, you'll find yourself going over budget in order to get the amount that you need. On the other hand, if you overprovision and invest in more resources, you'll end up wasting money.



When it comes down to it, many organizations find that they tend to overprovision their cloud resources. This is especially common when the team is used to maintaining hardware for an on-premises datacenter. In this case, it is often necessary to overprovision on hardware to meet future needs as your workloads and data continue to grow in the future. Adopting the cloud can be a culture shock in this respect, as it offers more flexibility to scale your cloud resources up and down as needed.

Overprovisioning is also very common for workloads that need high performance and are forced to buy more and more cloud resources in order to achieve the high level that they need.

In this ebook, we'll look more closely at overprovisioning and why companies end up overprovisioning on the cloud before examining ways that you can right-size your cloud resources to reduce unnecessary cloud spend.

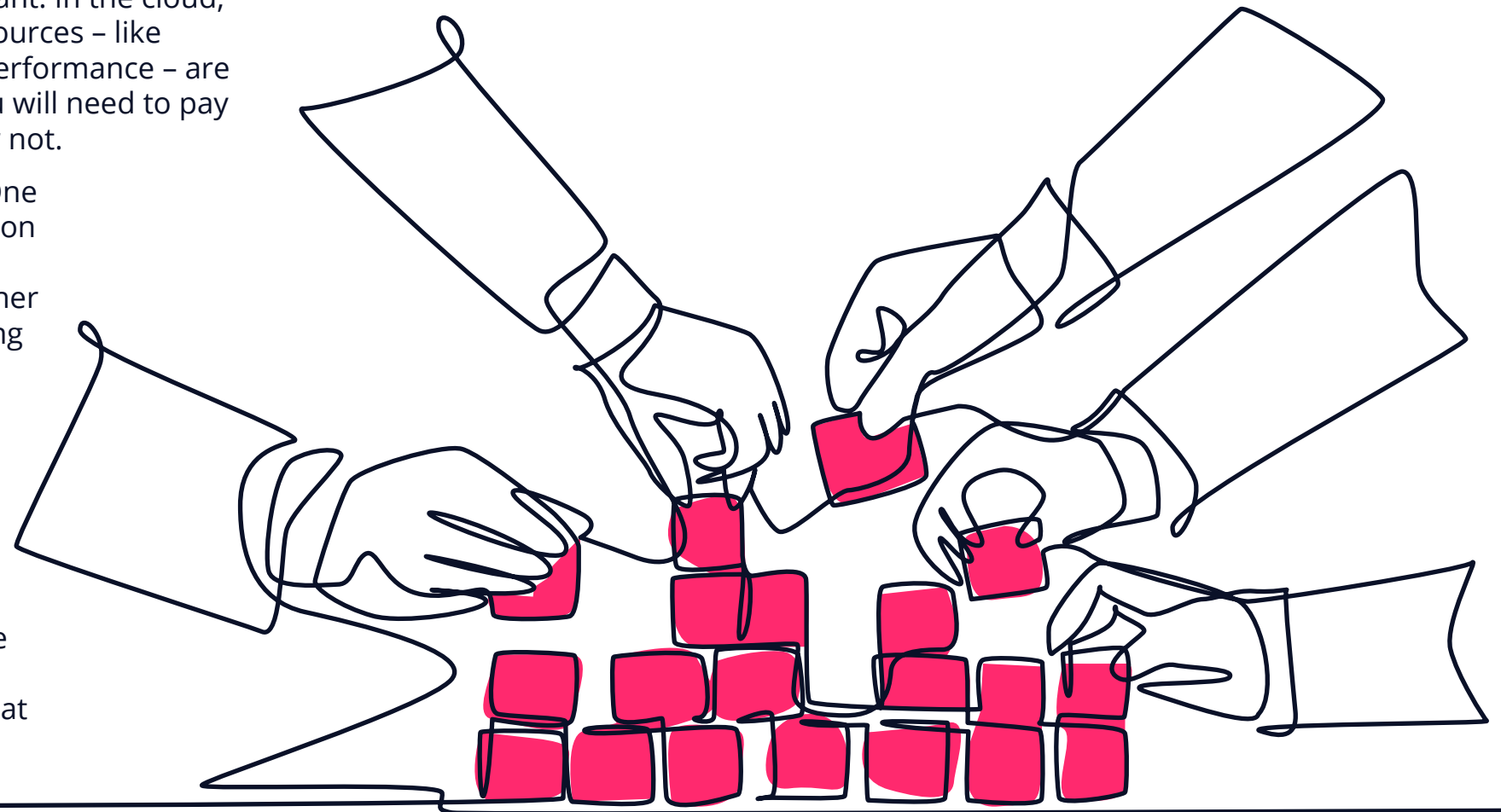


# Part 1: What is Overprovisioning and How Does It Happen?

Overprovisioning is the act of deploying and paying for resources you don't need in order to get enough of some other resource that you want. In the cloud, the biggest source of overprovisioning is the way that cloud resources – like compute, memory, network bandwidth, storage capacity, and performance – are packaged together. If you need more of one of these things, you will need to pay for more of the others, regardless of whether you need them or not.

There are many reasons an organization might overprovision. One is to absorb the highest end-user demands. But if the organization isn't careful, it can be all too easy to forget to turn off extra resources when they aren't needed -- thus creating waste. Another example of how this pain is felt is for organizations who are using the cloud for high-performing or mission-critical systems that require high performance. In order to get faster performance, they need to overprovision on other cloud resources, like capacity. Unfortunately, the cloud vendors set throttles on how much performance each of its customers can achieve.

This is due to the fact that the public cloud is a shared resource. And even with overprovisioning, cloud solutions are limited on their maximum capabilities and might not be able to meet the needs of all critical systems. So all the money in the world may not necessarily get you the ultra-fast performance that you require.

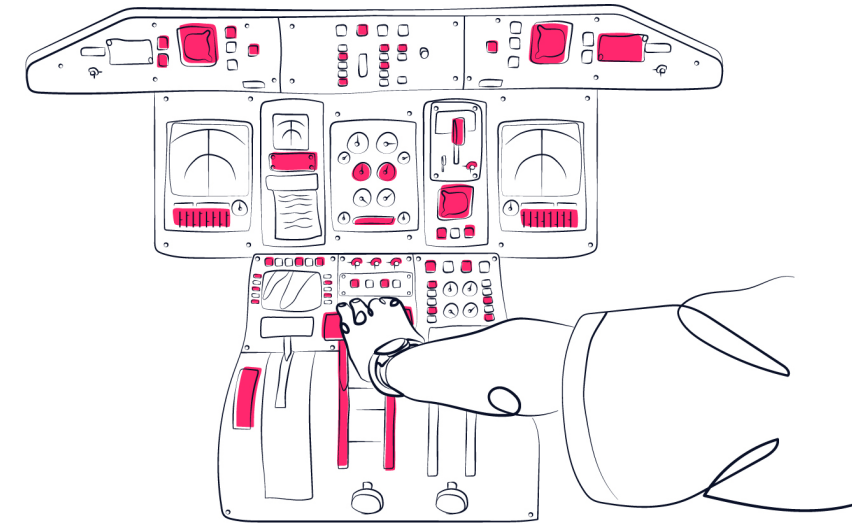
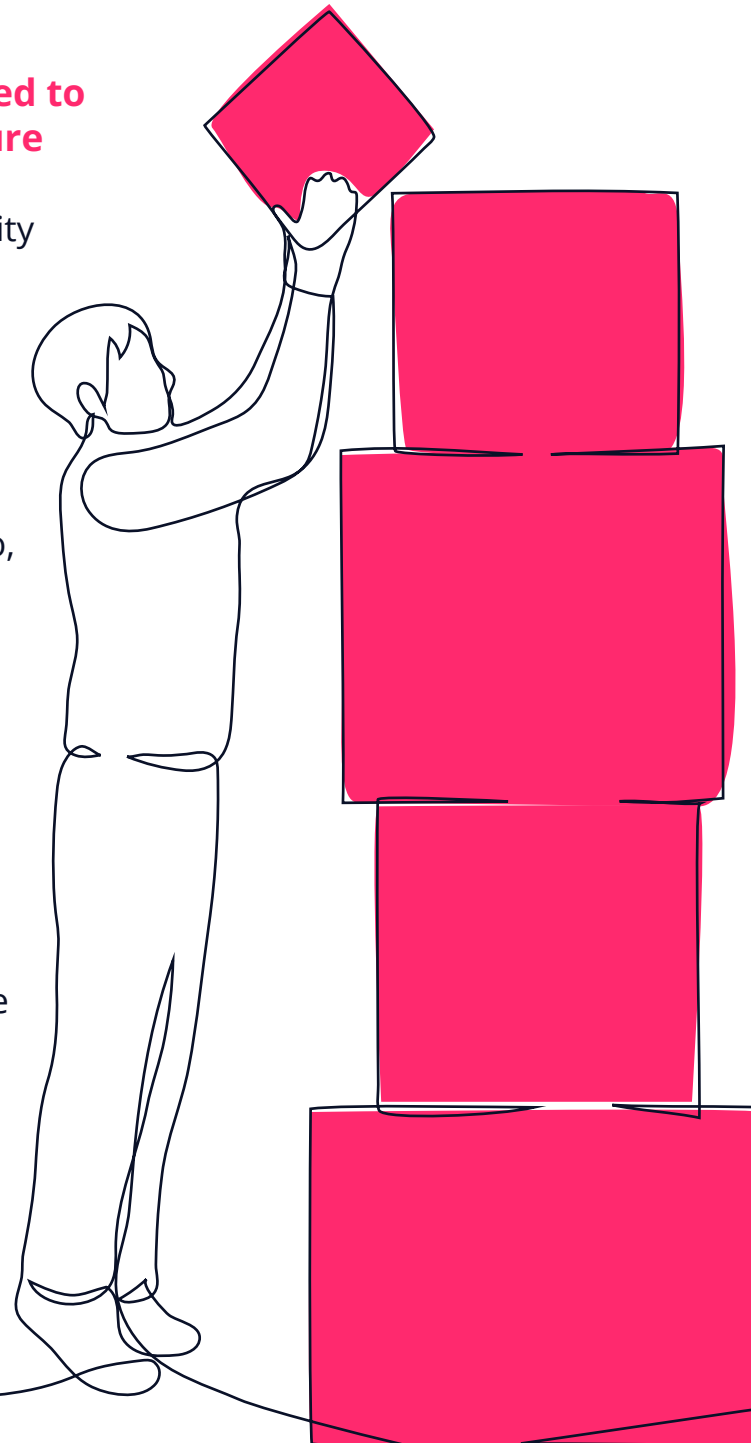


**Overprovisioning is not unique to the cloud. Organizations with on-premises datacenters are used to overprovisioning on hardware in order to meet future growth plans and needs.**

However, since the cloud is a self-service, on-demand, and virtually limitless in its scalability capabilities, it's much easier to overprovision. And with the major cloud providers presenting pricing in a "Price per hour" list view, the costs seem so miniscule, many users don't realize how much they have spent on overprovisioned resources until the bill comes in.

Additionally, some cloud storage solutions require extra capabilities from the virtual machines they are connected to, in order to get the full performance out of the connection. This leads to overprovisioning of the VMs, which increases the number of CPUs and, in turn, increases the cost of database software that is licensable by CPU cores, such as Oracle databases or Microsoft SQL Server. As such, you would need to buy additional database licenses for those additional cores (even though you don't need the extra CPU power). To avoid paying for these extra licenses, you can choose to restrict the VM instance so only a subset of the total available cores are used. This is similar to logical partitioning which is used on-prem. This drops the database licensing requirement – and cost – but it also restricts the compute performance.

The snag? You still have to pay the full cost of the VM, even though you are only using a fraction of its resources. So while you're avoiding unnecessary database licensing costs, you are still going to be billed month after month for compute cores that you literally cannot use.



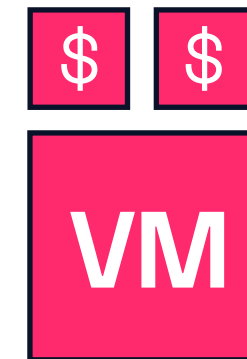
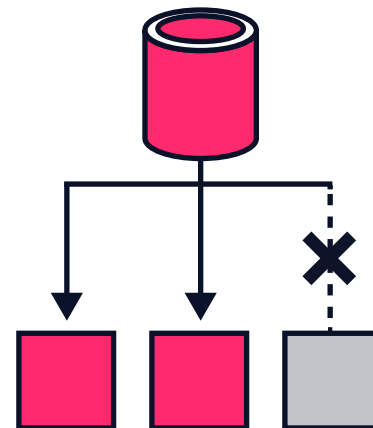
## Priceline's Story

Priceline is one company that has already taken steps to right-size its cloud resources in Google Cloud with the help of Silk. Priceline is the world leader in travel deals and had decided to move its workloads to Google Cloud in order to provide a better, faster customer experience without the maintaining the overhead of an on-premises datacenter. But staying within budget while getting the fastest performance possible was a major concern. With Silk, Priceline was able to cut back on overprovisioning while achieving its desired level of performance.

# Part 2: Right Sizing Your Cloud Resources

If you're currently overprovisioning your cloud resources, it's a good idea to take a hard look at what resources you are currently using and how you can reduce them to more accurately fit your cloud needs. The biggest benefit of right-sizing your cloud resources is reducing wasted expenditure. But going through the process of right-sizing can offer you a better understanding of what your cloud needs actually are. How exactly are your cloud resources being used? Are you spinning up resources for Dev/Test purposes and not turning them off when you're done – or are you spinning up multiple resources for the same purpose with no oversight? Are you using the best type of workloads and volumes for your needs? Right-sizing offers you better control over your cloud without compromising on price and performance.

**But how do you go about right-sizing your cloud environment? There are a few things you can do:**



**1. Identify usage patterns** – Monitor usage and performance for a certain length of time. This will allow you identify patterns in usage. Perhaps your load varies but the variation is predictable. Or maybe you have Dev/Test environments that are only being used during business hours and can be turned off on weekends and evenings.

**2. Turn Off Idle Instances** – If an instance is not being used for an extended period of time, it is time to shut it down. Of course, be sure to check with the owner of the instance to make sure there is no significant impact by turning this instance off (ie how difficult is it to recreate etc).

**3. Select the Right Instance Family** – Another way to right-size is to migrate to a different model in the instance family. If your max CPU and memory usage are much less than half over a month, you can cut the VM in half to save money. Or you can move to another instance family that is compatible in terms of virtualization type, network, and platform.

If you've tried all of the above ideas and are still unhappy with the overprovisioning of your cloud, there is still one more option available to you...

## Introducing Silk!

With the Silk Cloud DB Virtualization Platform, you can reduce the need to overprovision cloud resources and only pay for the additional resources you want. Silk allows you to dynamically scale performance or capacity as needed. The platform connects with compute VMs over a higher performance compute network, instead of the standard limited-capacity data network that the cloud providers use. This allows Silk to support the most performance-intensive workloads while eliminating the need to oversize compute VMs. In turn, this helps eliminate the need to overpay on unnecessary database license costs. Finally, Silk includes a package of enterprise data services, including zero-footprint snapshots, that make it easy to take copies of data without eating into your cloud capacity budget.

**Ready to cut down on unnecessary cloud resources?**

Visit [silk.us/cost](https://silk.us/cost) to learn more.

